

UNCLASSIFIED

AD 263 454

*Reproduced
by the*

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

CATALOG ASTIA 263 454

2-43-61-2

TECHNICAL REPORT

THRESHOLD COMPARISON OF
PHASE-LOCK, FREQUENCY-LOCK, AND
MAXIMUM-LIKELIHOOD TYPES
OF F M DISCRIMINATORS

263 454



JULY 1961

61-4-5
XF ROX

2-43-61-2

TECHNICAL REPORT

THRESHOLD COMPARISON OF
PHASE-LOCK, FREQUENCY-LOCK, AND
MAXIMUM-LIKELIHOOD TYPES
OF F M DISCRIMINATORS

by

J. J. SPILKER, JR.

JULY 1961

Lockheed

MISSILES and SPACE DIVISION

LOCKHEED AIRCRAFT CORPORATION • SUNNYVALE, CALIF

FOREWORD

The work reported herein was supported by Rome Air Development Center, Griffis Air Force Base, N. Y., under contract No. AF 30(602)-2232 and also as part of the Lockheed Missiles and Space Company General Research Program.

This paper was presented at the 1961 IRE WESCON conference, San Francisco, California, and has been submitted to the IRE for possible publication in the Proceedings or in one of the Transactions.

CONTENTS

	<u>Page</u>
ILLUSTRATIONS	iv
SUMMARY	1
INTRODUCTION	1
PHASE-LOCK DISCRIMINATOR	3
Description	3
Linearized Equivalent Circuit	3
Output SNR	4
Threshold Effect	5
Experimental Results	6
FREQUENCY-LOCK DISCRIMINATOR	7
Description	7
Linearized Equivalent	7
Threshold	9
Experimental Measurements	11
MAXIMUM-LIKELIHOOD FM DISCRIMINATOR	11
Description	11
COMPARISON OF THE DISCRIMINATORS	12
ACKNOWLEDGMENT	14
REFERENCES	14
APPENDIX A	14
Optimization of the Phase-lock Loop Transfer Function	14
APPENDIX B	15
Optimization of the Frequency-Lock Transfer Function	15

ILLUSTRATIONS

<u>Figure</u>		<u>Page</u>
1	Block Diagram of Phase-Lock FM Discriminator	2
2	Phase Correction Term for the Phase-Lock Loop	3
4	RC Loop Filter for the Phase-Lock Loop	4
5	Closed-Loop Attenuation Characteristics for the Phase-Lock Loop	4
6	Probability of Losing Lock for the Phase-Lock Loop	4
7	Measured Curve of Output vs. Input SNR for the Phase-Lock Discriminator	6
8	Measured Fractional Time out of Lock vs. Input SNR for the Phase-Lock Discriminator	6
9	Block Diagram of the Frequency-Lock FM Discriminator	7
10	Linearized Equivalent Circuit of the Frequency-Lock Loop	9
11	RLC Loop Filter for the Frequency-Lock Discriminator	9
12	Normalized Threshold Input $(\text{SNR})_T$ as a Function of the Deviation Ratio for the Frequency-Lock Discriminator	10
13	Threshold $(\text{SNR})_T$ as a Function of Deviation Ratio for the Frequency-Lock Discriminator	10
14	Measured Value of the Fractional Time Out of Lock vs. Input SNR for the Frequency-Lock Discriminator	11
15	Maximum-Likelihood FM Discriminator	11
16	An Approximate Computer Realization of the Maximum-Likelihood FM Discriminator	12
17	Threshold $(\text{SNR})_{in}$ as a Function of the Deviation Ratio for the Maximum-Likelihood Estimator	12
18	Comparison of the Threshold Input SNR for the Discriminators as a Function of D	13

THRESHOLD COMPARISON OF PHASE-LOCK, FREQUENCY-LOCK AND MAXIMUM-LIKELIHOOD TYPES OF FM DISCRIMINATORS *

J. J. Spilker, Jr.

Communications and Controls Research
Lockheed Missiles and Space Co.
Sunnyvale, California

Summary—In making a comparison of demodulation techniques for FM signals, one of the most important criteria for evaluation is the threshold input signal-to-noise ratio (SNR). It is at this value of input SNR that the output SNR begins to decrease, substantially more rapidly than a simple proportionality to the input SNR.

The objective of this paper is to present a consistent derivation of the threshold SNR for several major types of FM discriminators, namely, the phase-lock discriminator, the frequency-lock discriminator (FM discriminator with negative feedback), and the maximum-likelihood (a posteriori, most probable) computer. This last discriminator is of interest mainly because its operation provides a bound on the performance of other types.

The operation of the discriminators in the presence of gaussian interfering noise is described, and the causes of threshold are investigated. The probability of losing the "locked on" state of operation is computed. Curves of the threshold input SNR are plotted vs. the ratio of signal-to-modulation bandwidths, and these curves are compared for the various types of discriminators. Experimental confirmations of some of the theoretical results are given.

Introduction

In making a comparison of FM-demodulation techniques, one of the most important criteria for evaluation is the threshold input SNR. It is at this value of input SNR that the output SNR begins to decrease, substantially more rapidly than a simple proportionality to the input SNR. When a discriminator is operating well above its threshold and utilizes optimum filtering, it can be shown that the output SNR is linearly related to the input SNR by an expression that applies to almost any type of discriminator.

The main contribution of this paper is the derivation and comparison of the threshold input SNR for several major types of FM discriminators, namely, the phase-lock discriminator, the frequency-lock discriminator (FM discriminator with negative feedback), and the maximum-likelihood (a posteriori, most probable) computer. The operation and threshold behavior of these discriminators is analyzed in some detail, and the loop filters are optimized in order to minimize the value of the threshold input SNR.

*This work was supported in part by ARDC, Rome ADC, Griffiss AFB, N. Y.

Of the techniques discussed in this paper, an early version of the frequency-lock technique was the first to be presented in the literature. In 1939, J. G. Chaffee¹ described the use of feedback in an FM-demodulation operation. This discriminator did not require a limiter, but served both as an alternative to a limiter and as a method of reducing distortion.^{1,2,3,4} More recently, interest in the frequency-lock technique has been revived as a means for reducing the threshold input SNR.⁵ However, to the author's knowledge, no analysis of the threshold improvement is available in the literature.

The frequency-lock discriminator analyzed in this paper differs from the one used by Chaffee in that an ideal frequency discriminator is employed inside the feedback loop, whereas Chaffee's discriminator did not use a limiter and was amplitude sensitive. (An "Ideal" frequency discriminator is defined as a device where voltage output is proportional to the instantaneous frequency of its input. It can be accurately approximated by a wide-band limiter FM discriminator, or zero-crossing counting device.) This paper analyzes the performance and causes of threshold in the frequency-lock discriminator and optimizes the loop filters so as to minimize the threshold input SNR.

The phase-lock FM discriminator and other phase-locked circuits have received considerably more attention in the literature than has the frequency-lock discriminator. The first papers analyzing this technique seem to have appeared, about 1951.^{6,7,8,9} This set of papers discusses phase-locked circuits as applied to the problem of recovering the reference phase of a color TV signal. One of the more important applications of phase-locked circuits is in tracking systems, and numerous papers have appeared in the literature dealing with the design of phase-locked circuits optimized to track appropriate transients of phase.^{10,11,12} Other papers have presented new design and analysis techniques.^{13,14}

This treatment of the phase-lock discriminator differs from previous work in that it is aimed specifically at the problem of demodulating an FM signal which is modulated by analog intelligence having a low-pass spectrum. The purpose of the analysis is to determine the threshold behavior of a phase-lock discriminator which has been optimized (in a restricted sense) for this type of signal.

The maximum likelihood FM discriminator has been derived previously,^{15,16} and it has been shown that one of its forms uses the same configuration of elements as the phase-lock FM discriminator but that in this form it requires a nonrealizable loop filter. This

optimum discriminator can be realized, however, if the received data can be stored on tape (as is done in some telemetry systems) and processed by a special computer. This type of demodulation is not practical for most situations. Nevertheless, the study of its threshold performance provides a useful bound on the threshold behavior obtainable with other techniques.

In this paper, the threshold is evaluated on the basis of the expected percentage of time the discriminator will be in an unlocked state. When the probability of being in the unlocked state exceeds a certain prescribed small value under optimized conditions, the discriminator is assumed to have reached its threshold. Threshold is evaluated on the basis of attainable operation with a signal that is frequency modulated by a sine wave which has a maximum modulation frequency f_m and has a peak frequency deviation f_d . This modulation format is generally useful both for performance evaluation and for experimental verification. Stationary gaussian white noise is assumed as the interference.

The discriminator-transfer functions for both the phase-lock and the frequency-lock discriminators are optimized under the constraint that they be of second order or less. This constraint is imposed for practical reasons concerning both the analysis and the implementation. However, because of the stability requirements of these feedback circuits, little improvement is expected for higher-order transfer functions.

Phase-Lock FM Discriminator

Description

A block diagram of one form of the phase-lock FM discriminator is shown in Fig. 1. The received data are amplified by a band-pass filter, which has a bandwidth B_{pf} sufficient to pass the signal components relatively undistorted, and enter the multiplier of the phase-lock loop. When the discriminator is operating in its "locked-on" mode, the sine wave generated by the voltage-controlled oscillator (VCO) tracks the phase of the incoming signal; the multiplier output provides the phase correction information necessary to keep the VCO tracking properly. The loop filter which has a transfer function $F(p)$ removes much of the noise and high frequency components from the multiplier output and hence reduces their deleterious effect on the tracking operation.

Because of the nature of the phase-lock loop, the estimate of the signal phase obtained by the feedback system must be made in real time. A fixed delay T is usually tolerable, however, in obtaining the final frequency estimate at the receiver output. Therefore, post-detection filtering (transfer function $F_{pd}(p)$) is usually necessary to obtain the optimum frequency estimate when the T sec. delay is allowed and the filter constraints are thereby relaxed. The operation of the post-detection filter, however, does not affect the discriminator threshold and is of only minor concern in this paper.

It is common practice to precede the phase-lock loop with an AGC amplifier having constant power output, or a band-pass limiter (which also has constant power output). These elements restrict the dynamic range of the loop input and cause the loop gain to be dependent only upon the input SNR. If an ideal AGC amplifier is used,

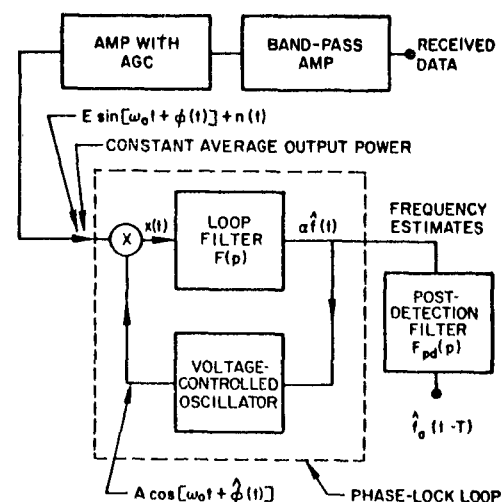


Fig. 1—Block diagram of phase-lock FM discriminator.

the statistics of the amplifier input and output are identical except for the appropriate gain factors, and, hence, the SNR at the loop input is the same as that at the AGC amplifier input. If a band-pass limiter is used instead, the input and output SNR differ only slightly and are related by ¹⁷ $(\text{SNR})_{\text{out}} = \alpha (\text{SNR})_{\text{in}}$ where $\pi/4 \leq \alpha \leq 2$. If $(\text{SNR})_{\text{in}} \gg 1$, we have $\alpha \approx 2$, and if $(\text{SNR})_{\text{in}} \ll 1$, we have $\alpha \approx \pi/4$. The noise statistics change considerably on passing through the band-pass limiter, however, and care must be used in applying this SNR relationship.

The ideal AGC amplifier is used in this analysis, although the results are shown to apply closely to the band-pass limiter as well. If the output power of the AGC amplifier is denoted by P_T , then the output signal amplitude E is related to the output noise power by $E^2/2 + P_n = P_T$. Since input and output SNR are identical for the amplifier, the signal amplitude is expressed by

$$E = \sqrt{2 P_T} / \sqrt{1 + r} \quad (1)$$

where $r \triangleq 1/(\text{SNR})_{\text{in}}$.

Linearized Equivalent Circuit

The first step in the analysis of the phase-lock loop is to compute the low-frequency terms that are present in the multiplier output. For this purpose, let us express the signal component at the output of the AGC amplifier as

$$E \sin [\omega_0 t + \phi(t)]$$

where

$$\phi(t) \triangleq \int_0^t f(\omega) d\omega$$

the phase modulation produced by a frequency modulating message waveform $f(t)$. Define the VCO output as

$$\hat{s}(t) = A \cos [\omega_0 t + \hat{\phi}(t)]$$

and the noise at the multiplier input as

$$n(t) = N(t) \sin [\omega_0 t + \psi(t)]$$

Then, by referring to Fig. 1, the multiplier output $x(t)$ can be expressed as

$$2x(t)/A = E \sin [\phi(t) - \hat{\phi}(t)] + N(t) \sin [\psi(t) - \hat{\phi}(t)]$$

$$= E [\sin \epsilon(t) + \phi_n(t)]$$

where terms in the frequency range $2f_0$ have been neglected, the phase error is denoted by $\epsilon \triangleq \phi - \hat{\phi}$, and

$$\phi_n(t) \triangleq [N(t)/E] \sin [\psi(t) - \hat{\phi}(t)].$$

The first term in this expression represents a phase correction term, and the second represents interference.

When the discriminator is operating above threshold, the phase of the VCO is closely tracking the phase of the signal, i.e., $\hat{\phi}(t) \approx \phi(t)$. In this paper, the signal and noise components are considered to be statistically independent. Thus, when the discriminator is operating above threshold, the power spectral density of $\phi_n(t)$ can be expressed approximately as the low-frequency portion of the convolved VCO and noise spectra; i.e.,

$$E^2 G_{\phi_n}(f) \sim G_S(f) * G_N(f),$$

where $G_{\phi_n}(f)$, $G_S(f)$, and $G_N(f)$, represent the power spectra of $\phi_n(t)$, $\hat{s}(t)$, and $n(t)$, respectively.

The phase correction term $E \sin \epsilon$ can be approximated by the piecewise linear curve shown in Fig. 2. In the region of phase error, $|\epsilon| < 1$ the multiplier output is approximated by

$$2x(t)/A \approx E [\epsilon(t) + \phi_n(t)] = E [\phi_c(t) - \hat{\phi}(t)] \quad (2)$$

where $\phi_c \triangleq \phi + \phi_n$ and can be termed the equivalent phase input. This last expression leads to the linearized equivalent circuit for the phase-lock loop shown in Fig. 3.

This circuit is equivalent to the actual nonlinear circuit in that the loop-filter input is identical with that given by (2), and hence the outputs of the two circuits are the same. Notice that in this context, the term $\phi_n(t)$ can be considered the equivalent input phase noise, it can be shown to be approximately the same as the phase noise term that would appear at the output of a bandpass limiter at high input SNR (>10). Hence, the use of a band-

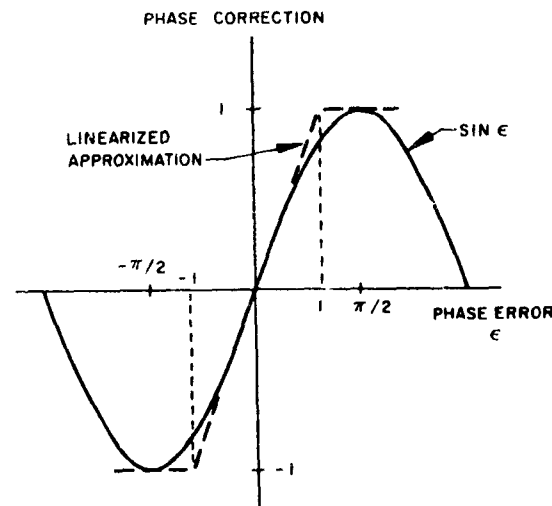


Fig. 2—Phase correction term for the phase-lock loop.

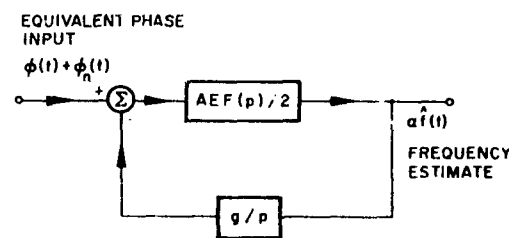


Fig. 3 Linearized equivalent circuit for the phase-lock loop.

pass limiter at high-input SNR is not expected to perform differently than with the AGC amplifier. At low SNR, the output SNR of the bandpass limiter is $\pi/4$ times the input SNR, or only 1.0 db below the input SNR. Hence, the results computed on the basis of an AGC amplifier should correspond closely to those for a bandpass limiter.

The transfer function of the linearized circuit (equivalent phase input-to-frequency estimate) is

$$\frac{\hat{f}(p)}{\Phi_c(p)} = \frac{g_c F(p)}{1 + g_c F(p)/p}$$

where $\hat{f}(p)$ and $\Phi_c(p)$ are the Laplace transforms of $\hat{f}(t)$ and $\phi_c(t)$, respectively, and we define $g_c = gAE/2$, the equivalent gain. Hence, the phase-transfer function (equivalent phase input-to-phase estimate, transfer function and the transfer function of equivalent frequency-to-frequency estimate) is expressed by

$$H(p) \triangleq \frac{\hat{\phi}(p)}{\Phi_c(p)} = \frac{g_c F(p)/p}{1 + g_c F(p)/p} \quad (3)$$

where $\hat{\phi}(p)$ is the Fourier transform of $\hat{\phi}(t)$.

Two types of phase-error components are present in the loop—transient error and noise error. The transient phase error is defined as the error in ϕ which is dependent only upon the input signal; it has the transform $E_t(p)$

$$E_t(p) = \Phi(p) [1 - H(p)] \quad (4)$$

The phase-noise error is defined as the error in the phase estimate which is dependent upon the input noise, and it has the transform $E_n(p)$

$$E_n(p) = \Phi_n(p) H(p) \quad (5)$$

Both (4) and (5), of course, are based upon the assumption that the linearized analysis is a good approximation. These expressions are used in the next section to evaluate the discriminator performance.

The dynamic (frequency-tracking) range of the discriminator Δf is a fundamental parameter of the loop and is defined as the maximum steady-state frequency shift that can be produced by the VCO. The maximum average multiplier output occurs with $\epsilon = \pi/2$ and has the value $x_{\max} = EA/2$. Notice that on the basis of the linear model, this maximum frequency shift occurs for $\epsilon = 1$. The corresponding maximum VCO frequency shift is $\Delta f = EA_g F(0)/2 = g_e F(0)$.

The product $g_e F(0)$ represents the dc loop gain, and it is convenient to define the quantity $g_o = g_e F(0)$ for use in the next section.

Threshold Effect

In the piecewise linear approximation to the phase correction term of the actual phase-lock loop, the correction is linear with respect to phase error for $|\epsilon| < 1$. If $|\epsilon|$ should exceed 1, the linear behavior no longer holds—the phase estimate becomes highly distorted, and the loop can be unstable. In this paper, the discriminator is considered to reach its threshold when the input SNR decreases to the point where the probability of having $|\epsilon| > 1$, computed on the basis of the linearized analysis, exceeds a prescribed small probability. More specifically, threshold is defined to occur when the rms value of the total phase error $\sigma_T = 1/2$. A certain degree of arbitrariness is inherent in any definition of threshold, and this one is no exception. It is difficult to specify at just what probability of losing lock, or how far below the linear behavior of the output SNR curve, threshold should be said to occur. However, some measure of the arbitrariness is removed in a later paragraph where the probability of losing lock is computed as a function of the input SNR. The curve of probability of losing lock vs. input SNR which is obtained gives insight into the variation of the discriminator performance as the input SNR varies slightly above or below its threshold value. In types of FM discriminators which track the phase or frequency of the input signal the percent time out of lock is perhaps a more meaningful measure of performance than the output SNR because near threshold large errors in the estimate occur in bursts. In the time intervals between bursts, the output noise level is relatively small, and the accuracy in the message estimate is relatively good.

Threshold is evaluated on the basis of a signal which has a sinusoidal phase modulation of the form

$$\phi(t) = (f_d/f_a) \cos \omega_a t$$

where $f_a \leq f_m$ the maximum modulation of frequency and f_d is the peak frequency deviation. The transient-phase error with this modulation is a sinusoid of frequency f_a having a peak value

$$\epsilon_t = (f_d/f_a) |1 - H(j\omega_a)| \quad (6)$$

The value of peak transient error used for the threshold determination is the maximum value of

$$\epsilon_t \text{ for } f_a \leq f_m.$$

The phase noise has a mean square value

$$\sigma_n^2 = \int_{-\infty}^{\infty} G_{\phi_n}(f) |H(j\omega_a)|^2 df \quad (7)$$

Thus the total rms phase error is

$$\sigma_T = (\epsilon_t^2/2 + \sigma_n^2)^{1/2}$$

can be computed from (6) and (7); this quantity determines the threshold.

The closed-loop transfer function considered in this paper is constrained to have the form

$$H(p) = \frac{1 + ap}{1 + b_1 p + b_2 p^2} \quad (8)$$

Appendix A shows that the approximately optimum form of this transfer function for large deviation ratio is

$$H(p) = \frac{1 + ap}{1 + ap + a^2 p^2} \quad (9)$$

This transfer function is approximately optimum for the deviation ratios of main interest in this paper

$$D \triangleq f_d/f_m > 1.$$

The loop filter required for the realization of this transfer function is the RC filter depicted in Fig. 4. The closed-loop attenuation characteristic $|H(j\omega)|$ is plotted in Fig. 5. The peak transient error occurs for $f_a = f_m$ with this filter.

It is shown in Appendix A that the input SNR is related to the peak transient error and the phase noise by the expression

$$(SNR)_{in} = \frac{1}{\sigma_n^2 \epsilon_t^2} \pi D^{1/2} f_m \quad (10)$$

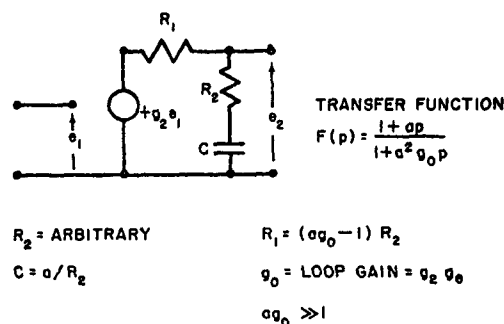


Fig. 4-RC loop filter for the phase-lock loop.

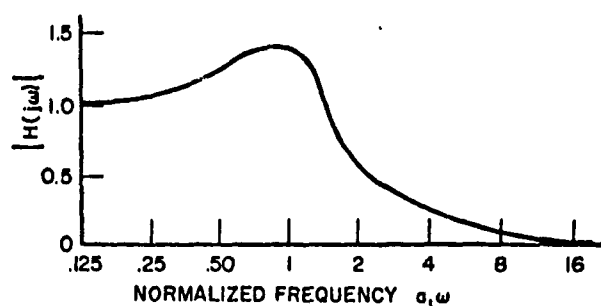


Fig. 5-Closed-loop attenuation characteristics for the phase-lock loop.

The value of a used is $a = c_t^{1/2} / \omega_m D^{1/2}$. The quantity B_{rf} denotes the equivalent noise bandwidth of the RF amplifier.

If this expression is minimized with respect to ϵ_t for a prescribed rms total error, it becomes

$$(\text{SNR})_{in} = \frac{1.57}{\sigma_T^{5/2}} \frac{\pi D^{1/2} f_m}{1/2}$$

for $\epsilon_t = \sqrt{2/5} \sigma_T$, $\sigma_n = \sqrt{4/5} \sigma_T$. Thus the threshold SNR is given by

$$(\text{SNR})_{inT} = 8.9 \frac{\pi D^{1/2} f_m}{B_{rf}} \quad (11)$$

Consider that the RF filter is single-tuned and has a 3 db bandwidth $2(D+1)f_m$, then the equivalent noise bandwidth is $B_{rf} = \pi(D+1)f_m$. The threshold expression (11) then becomes

$$(\text{SNR})_{inT} = 8.9 D^{1/2} / (D+1)$$

Notice that for large deviation ratios, $D \gg 1$, the threshold input SNR is proportional to $D^{-1/2}$. Equation (12) is discussed further in the final section of the paper.

The probability that the phase error will exceed the threshold error $|\epsilon| = 1$ based on the linearized analysis, can be computed as a function of the input SNR. If the phase noise in the estimate is assumed to be approximately gaussian, then the total phase error is the sum of a sinusoid plus a gaussian term; the statistics of this sum are available from the literature¹⁸. At threshold, the probability of exceeding $|\epsilon| = 1$ is about 0.043. The probability of losing the locked-on state is plotted in Fig. 6 under the assumption of a fixed linearized closed-loop transfer function $H(p)$.

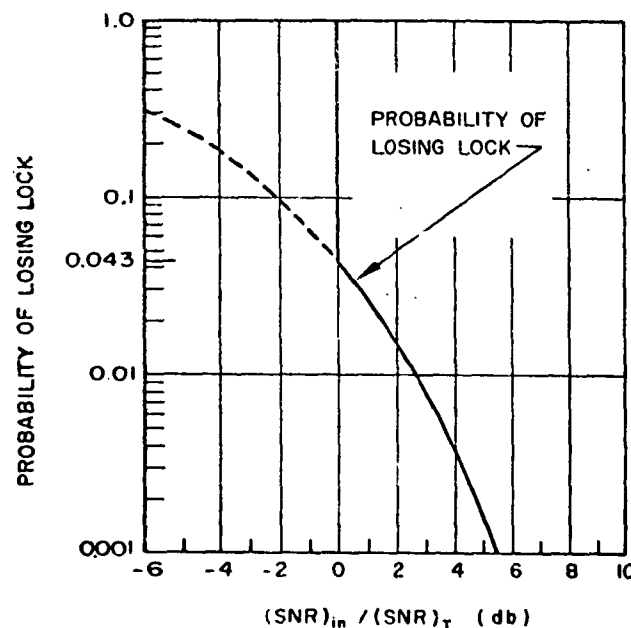


Fig. 6-Probability of losing lock for the phase-lock loop.

Because of nonlinear effects, the probability of losing lock is assumed to be essentially unity for input SNR below the threshold value. When the discriminator loses lock, it must go through a rather complex lock-in operation before linear operation is resumed.⁷

Output SNR

When the post-detection filter is properly optimized at high-input SNR for a low-pass modulation $f(t)$, the combined transfer function of the phase-lock loop and the post-detection filter has a constant attenuation characteristic and a linear phase characteristic for $|f| \leq f_m$. The filter greatly attenuates frequency components above the modulation pass-band $|f| > f_m$. When the probability of losing lock is small, the input

and output SNR of the complete discriminator can be shown to be related by the expression

$$(\text{SNR})_{\text{out}} = \frac{3}{2} D^2 \frac{B_{\text{rf}}}{f_m} (\text{SNR})_{\text{in}} \quad (13)$$

A sinusoidal frequency modulation having a peak deviation $f_d = Df_m$, has been assumed for this calculation.

Experimental Results

A phase-lock FM discriminator has been constructed with filter parameters based on the theory presented herein; it has the following modulation characteristics:

Message bandwidth	$f_m = 2\text{kc}$
Peak frequency deviation	$f_d = 10.5\text{kc}$
Deviation ratio	$D = 5.25$
Equivalent noise bandwidth	$B_{\text{rf}} = (\pi/2)60\text{kc} = 94\text{kc}$

This experimental discriminator utilized a band-pass limiter rather than an AGC amplifier. However, it is not expected to alter the discriminator performance significantly from that produced by the AGC amplifier, because the main region of interest is for input SNR ≥ 1 . A measured curve of output vs. input SNR for the discriminator is plotted in Fig. 7. A theoretical curve is

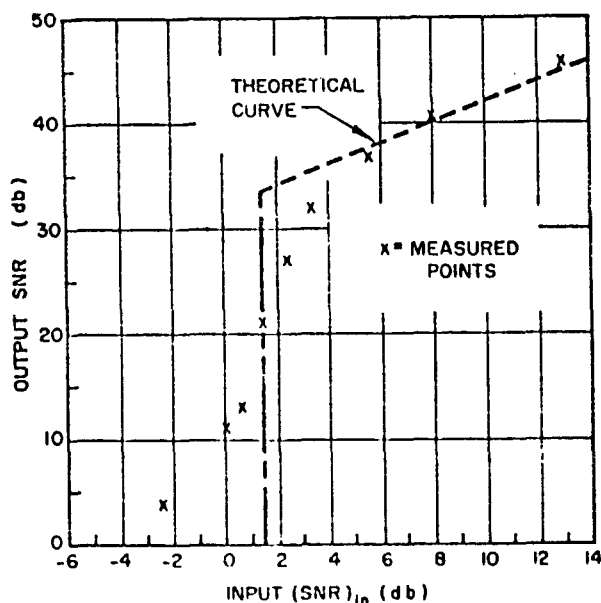


Fig. 7 Measured curve of output vs. input SNR for the phase-lock discriminator.

plotted in the same figure which obeys (13) for input SNR above the threshold value and drops to zero at and below threshold. The threshold input SNR computed from (11) is 1.4 db. There is good agreement between the measured and the theoretical curve. It can be seen, however, that the output SNR decreases noticeably below the linear curve at input SNR, 2 db above the threshold value; at threshold, the output SNR decreases to + 21 db.

A measured curve of the fractional time out of lock* is shown in Fig. 8 along with a theoretical curve based on

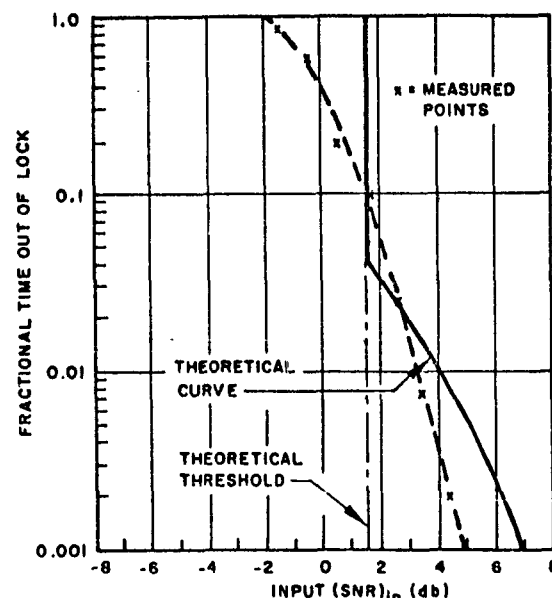


Fig. 8—Measured fractional time out of lock vs. input SNR for the phase-lock discriminator.

Fig. 6. (Note that the fractional out-of-lock scale is logarithmic.) The agreement between these curves is quite good. The measured value of the fractional time out of lock is 0.043 db for an input SNR (threshold) of 2.2 db, which is only 0.7 db above the predicted threshold. The measured fractional time out of lock curve changes at a rate greater than that given by the steeper linearized analysis under the assumption of gaussian noise, but does not, of course, have a sudden discontinuity at the threshold. A partial explanation for the steeper slope can be found by noting that these measurements were made using a bandpass limiter and, as a result, the phase-noise statistics deviate from the gaussian statistics that have been assumed. If the limiter effect is taken into account in computing the curve in Fig. 6, the theoretical curve would be considerably steeper in the region above threshold and agree more closely with the measured data of Fig. 8.

*The fractional time out of lock is determined by measuring the fractional time in which large bursts of noise appear in the loop output. These noise bursts are at least an order of magnitude larger than the noise level between bursts and hence, are readily separated from this ambient noise.

Frequency-Lock Discriminator

Description

The frequency-lock FM discriminator is shown in block diagram form in Fig. 9. As the figures shows, the in-

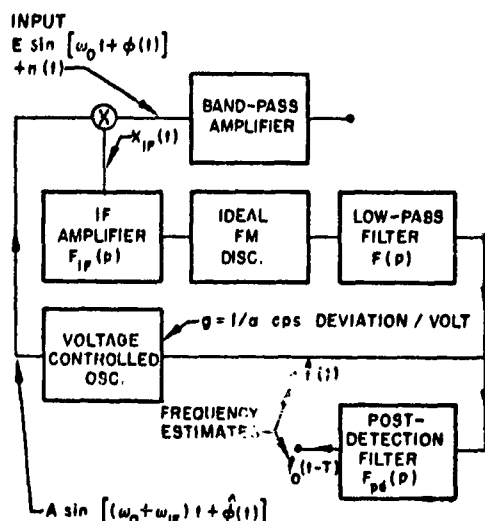


Fig. 9-Block diagram of the frequency-lock FM discriminator.

coming signal plus additive gaussian white noise is first amplified by a band-pass amplifier of bandwidth B_{rf} sufficient to pass the signal with low distortion, and is then mixed with a sinusoid generated by the VCO. When the discriminator is tracking properly, the instantaneous frequency of the VCO output closely tracks the instantaneous frequency of the signal with a fixed IF frequency off-set, f_{IF} . If the deviation ratio is greater than unity, the condition of greater interest in this paper, then the residual frequency deviation of the signal after mixing can be considerably less than that of the RF transmission. Because of this reduction in the deviation, an IF amplifier bandwidth less than the RF bandwidth can be used, and much of the noise can be removed before the signal reaches the FM discriminator inside the loop.

The ideal FM discriminator used in this loop generates a voltage output proportional to the instantaneous frequency of its input. This discriminator provides a correction voltage which is dependent upon the frequency tracking error. The loop filter $F(p)$ serves to eliminate much of the frequency noise from the discriminator while it passes the modulation components with low distortion; the filter output is the estimate $\hat{f}(t)$ of the modulation. A post-detection filter $F_{pd}(p)$ is used to obtain the optimum estimate of the modulation with a time delay T , i.e., $\hat{f}_o(t - T)$. As was true with the phase-lock loop, the post-detection filter does not affect the discriminator threshold.

Linearized Equivalent of the Frequency-Lock Discriminator

The analysis of the frequency-lock discriminator begins with the computation of the components in the mixer output. It is convenient for this purpose, to express the signal and VCO output waveforms in the forms:

$$s(t) = E \sin [\omega_0 t + \phi(t)]$$

and

$$\hat{s}(t) = A \sin [(\omega_0 + \omega_{IF})t + \hat{\phi}(t)]$$

where

$$\phi(t) = \int_0^t f(\mu) d\mu$$

and

$$\hat{\phi}(t) = \int_0^t \hat{f}(\mu) d\mu.$$

If we again represent the noise as

$$n(t) = N(t) \sin [\omega_0 t + \psi(t)],$$

then by referring to Fig. 9, the mixer output $x_{IF}(t)$ can be expressed by

$$\begin{aligned} 2x_{IF}(t)/AE &= \cos [\omega_{IF}t + \phi(t) - \hat{\phi}(t)] \\ &+ \frac{N(t)}{E} \cos [\omega_{IF}t + \psi(t) - \hat{\phi}(t)] \\ &= \cos [\omega_{IF}t + \phi(t) - \hat{\phi}(t)] + n_2(t) \end{aligned}$$

where

$$n_2(t) \triangleq [N(t)/E] \cos [\omega_{IF}t + \psi(t) - \hat{\phi}(t)].$$

Terms in the frequency range $f_0 \pm f_{IF}$ are neglected. The first term in this expression is the frequency correction term; the second is an interference term caused by the noise input. The residual frequency deviation in the correction term can be defined as

$$f_T(t) \triangleq f(t) - \hat{f}(t).$$

Consider that the noise $n(t)$ is gaussian; then the noise term in the mixer output is also gaussian. When the discriminator is operating above threshold the frequency modulation and the estimate of this modulation are approximately equal. Hence the spectrum of $n_2(t)$ is approximately equal to the low frequency portion of the convolution $G_{\hat{s}}(f) * G_n(f)$.

The mixer output is passed through an IF filter having a symmetrical (high Q) single-tuned response and a 3-db bandwidth B_{IF} . The transfer function of this filter is

$$F_{IF}(p) = \frac{1}{1 + (p - j\omega_{IF})/\pi B_{IF}} + \frac{1}{1 + (p + j\omega_{IF})/\pi B_{IF}}$$

This transfer function can also be expressed in the forms

$$F_{IF}(j\omega) \triangleq |F_{IF}(j\omega)| e^{j\theta(j\omega)} \\ = F_a(j\omega - j\omega_{IF}) + F_a(j\omega + j\omega_{IF}) \quad (14)$$

The function $F_a(j\omega)$ is the low-pass equivalent of the IF filter.

The signal component at the IF output is distorted by an amount which is dependent upon a complex relationship between peak value of the residual frequency deviation f_{rp} , the maximum modulation frequency f_m , and the IF bandwidth B_{IF} . Baghdady¹⁹, however, has shown that a quasi-stationary response is obtained with a single-tuned high Q filter if $4f_m f_{rp} \ll B_{IF}^2$. In the quasi-stationary approximation, a sinusoid of instantaneous angular frequency $\omega_i(t)$ on passing through a filter having a transfer function $|F_{IF}(j\omega)| \exp j\theta(j\omega)$ is amplitude distorted by $|F_{IF}(j\omega_i(t))|$ and phase distorted by $\theta[j\omega_i(t)]$. In this treatment of the frequency-lock discriminator the IF bandwidth is considered to be large enough so that quasi-stationary analysis is a good approximation, and the signal component in the IF output is represented by

$$\frac{1}{2} AE |F_{IF}(j\omega_r(t))| \cos \left[\omega_{IF} t + \int_0^t \omega_r [\mu - \theta'(j\omega_{IF})] d\mu \right] \quad (15)$$

where $\omega_r(t) \triangleq 2\pi f_r(t)$ is the residual angular modulation, and $\theta'(j\omega_{IF})$ is the group delay of the filter, $\theta'(j\omega) \triangleq d\theta(j\omega)/d\omega$. Therefore, the main effect on the residual frequency modulation is a delay of $\theta'(j\omega_{IF})$ sec.

Denote the IF output noise component as

$$n_{IF}(t) = N_{IF}(t) \sin [\omega_{IF} t + \psi_{IF}(t)].$$

If the residual phase modulation of the signal is denoted by $\phi_r(t)$, then the frequency noise component in the discriminator output is the derivative of the phase noise $\phi_n(t)$. By the use of a phasor diagram this phase noise can be evaluated as

$$\phi_n = \tan^{-1} [N_{IF}(t)/E] \sin [\psi_{IF}(t) - \phi_r(t)] \\ \approx [N_{IF}(t)/E] \sin [\psi_{IF}(t) - \phi_r(t)] \quad (16)$$

$$\text{if } |N_{IF}(t)| \ll E$$

At SNR above threshold, the assumption that $|N_{IF}(t)| \ll E$ is a valid one. It can be seen that if the residual modulation, $f_r(t) = d\phi_r(t)/dt$, is small,

then the Fourier transform of $\phi_n(t)$, denoted as $\Phi_n(j\omega)$ is approximately equal to

$$\Phi_n(j\omega) \approx N_a(j\omega) F_a(j\omega) \quad (16a)$$

where $N_a(j\omega)$ is the Fourier transform of

$$N_a(t) = [N(t)/E] \cos [\psi(t) - \hat{\phi}(t)].$$

This last expression shows the phase noise to be the same as a heterodyned version of the mixer output noise, namely, $N_a(t)$, which has been filtered by a low-pass equivalent, $F_a(j\omega)$, of the IF filter. However, the assumption that this statement is based upon, namely, $f_r(t) \approx 0$, is not completely accurate.

Nevertheless, it can be seen that when the discriminator is locked on, the signal component in the mixer output, $\cos [\omega_{IF} t + \phi_r(t)]$, is of small enough bandwidth to pass the IF relatively undistorted. Hence, the additional phase modulation $\phi_r(t)$, which appears in (16), is not large enough to change greatly the power spectrum of $\phi_n(t)$ from that which would be obtained from (16a).

Under the conditions stated above, therefore, the output of the ideal discriminator is

$$k \left\{ f_r [t - \theta'(j\omega_{IF})] + \frac{1}{2\pi} \frac{d\phi_n(t)}{dt} \right\}$$

where k is the gain of the discriminator.

The approximate Fourier transform of this term is

$$k \left[\mathcal{F}(j\omega) \hat{\mathcal{F}}(j\omega) + j f N_a(j\omega) \right] F_a(j\omega) \quad (16a)$$

where $\mathcal{F}(j\omega)$, $\hat{\mathcal{F}}(j\omega)$ are the Fourier transforms of $f(t)$ and $\hat{f}(t)$, respectively. The approximation is made that the dominant effect of $F_a(j\omega)$ on the residual frequency deviation term is to cause the delay $\theta'(j\omega_{IF})$ as in (15).

Thus by reference to Fig. 9 and (16a) it can be seen that the Fourier transform of the frequency estimate is approximately equal to

$$\alpha \hat{\mathcal{F}}(j\omega) \approx F_a(j\omega) F(j\omega) [\mathcal{F}(j\omega) - \hat{\mathcal{F}}(j\omega) - \mathcal{F}_{ne}(j\omega)]$$

where $\mathcal{F}_{ne}(j\omega) \triangleq j \mathcal{L} N_a(j\omega)$ is the transform of $f_{ne}(t)$, the equivalent frequency noise input. Under conditions of locked-on operation, the frequency estimate obtained by the frequency-lock discriminator can be seen to be the same as that provided by the linearized equivalent circuit shown in Fig. 10. The closed-loop transfer function of this circuit is $\omega H(j\omega) = \alpha \mathcal{F}(j\omega) \hat{\mathcal{F}}(j\omega)$ where

$$H(j\omega) \triangleq \frac{\hat{\mathcal{F}}(j\omega)}{\mathcal{F}(j\omega)} = \frac{g_c F_a(j\omega) F(j\omega)}{1 + g_c F_a(j\omega) F(j\omega)} \quad (17)$$

and the equivalent gain $g_c \triangleq kg$

As was true with the phase-lock loop, two types of error components are produced in the estimation process — transient error and noise error. The transient frequency error, which results from a

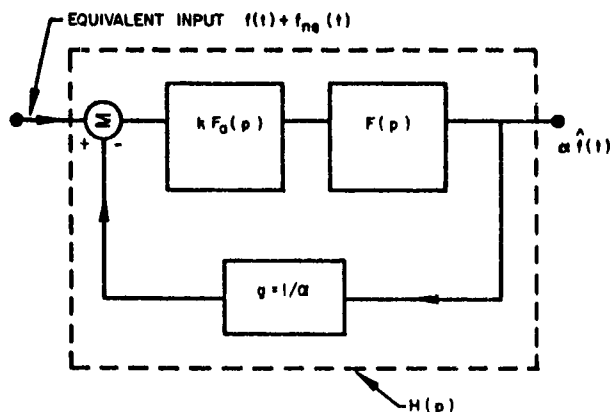


Fig. 10—Linearized equivalent circuit of the frequency-lock loop.

frequency modulation having the transform $\mathcal{F}(j\omega)$, has the transform $\mathcal{F}_t(j\omega)$.

$$\mathcal{F}_t(j\omega) = \mathcal{F}(j\omega) [1 - H(j\omega)] \quad (17a)$$

The output frequency noise has the transform

$$\mathcal{F}_n(j\omega) = \mathcal{F}_{ne}(j\omega) H(j\omega) \quad (17b)$$

Threshold

In an "ideal" FM discriminator,* the output SNR begins to decrease, substantially more rapidly than the first power of the input SNR when the input SNR goes as low as +10 db. In the frequency-lock discriminator, the threshold effect occurs when the SNR at the input to the ideal discriminator decreases below 10 db, and signal suppression effects begin to be of significance.²⁰ This IF SNR can decrease to its threshold value from two causes: (1) The noise power at the input to the frequency-lock discriminator can increase relative to the signal power. (2) The residual frequency deviation in the mixer output can increase to the point where the signal component falls outside the IF pass-band and is attenuated relative to the noise and otherwise distorted. The objective of the optimization that is performed in this paper is to adjust the IF filter bandwidth, the loop gain, the loop-filter transfer function so that these two causes of threshold occur at as low an input SNR as possible.

The optimization is to be performed by minimizing the rms total frequency error σ_{f_T} by properly choosing the loop transfer function coefficients. The IF-filter 3-db bandwidth B_{IF} is then given the smallest possible value to permit significant distortion of the signal component (IF frequency correction component) to begin at threshold. The signal component in the IF is considered

*Broad-bandwidth pulse counting or Foster-Seely types of FM discriminators are good approximations to an ideal FM discriminator

to require a minimum bandwidth of approximately $2(f_m + \sqrt{2} \sigma_{f_T})$ for low distortion when the maximum modulation frequency is f_m . Notice that for sinusoidal frequency error, the peak error deviation is $\sqrt{2}$ times the rms frequency error. Threshold is said to occur when the IF bandwidth has this value, and the IF output SNR decreases to 10 db.

The frequency modulation of the signal is assumed to have the form $f_d \cos \omega_a t$ where $\omega_a \approx \omega_m$. When the discriminator is locked-on and the linearized analysis (17a) is accurate the peak transient error f_{tp} has the value

$$f_{tp} = f_d |1 - H(j\omega)| \quad (18)$$

The rms frequency noise error can be found from (17b) to be

$$\sigma_{f_n}^2 = \int_{-\infty}^{\infty} G_{f_{ne}}(f) |H(j\omega)|^2 df \quad (19)$$

where $G_{f_{ne}}(f)$ is the power spectral density of $f_{ne}(t)$, the equivalent frequency noise input.

The closed-loop linearized equivalent transfer function of the frequency-lock discriminator is constrained to have the form

$$H(p) = g/(1 + b_1 p + b_2 p^2)$$

Appendix B shows that the approximately optimum form of this transfer function is

$$H(p) = \frac{g}{1 + bp + bp^2/\omega_m}$$

The loop filter used to obtain this transfer function is shown in Fig. 11. The peak transient frequency error (18), now occurs when the modulation frequency is at

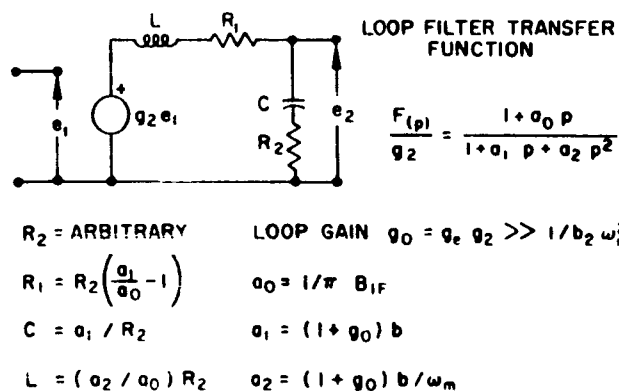


Fig. 11—RLC loop filter for the frequency-lock discriminator

its maximum f_m . With this transfer function, the peak transient and noise frequency errors are related to the input SNR by the equation

$$(\text{SNR}) = \frac{\pi^3 f_m^2}{B_{rf} \sigma_{f_n}^2} \quad (21a)$$

This relationship is obtained from (19) by setting

$$b = f_{tp}/2^{1/2} f_{tp}^2 \omega_m,$$

the value obtained by solving (18) for b with $f_a = f_m$. Now, for a given $(\text{SNR})_{in}$, the rms total error

$$\sigma_{fT} = \sqrt{f_{tp}^2/2 + \sigma_{f_n}^2}$$

is minimized with respect to f_{tp} by setting

$$f_{tp}/2^{1/2} = \sigma_{f_n} = \sigma_{fT}/2^{1/2}.$$

By using (21a) the rms total frequency error then can be expressed by (see Appendix B)

$$\sigma_{fT} = f_m D^{1/2} \left[\frac{2\pi f_m}{B_{rf}(\text{SNR})_{in}} \right]^{1/4} \quad (22)$$

The minimum IF bandwidth consistent with this frequency error is then

$$B_{IF} = 2(f_m + \sqrt{2} \sigma_{fT})$$

$$= 2f_m \left\{ 1 + (2D)^{1/2} \left[\frac{2\pi f_m}{B_{rf}(\text{SNR})_T} \right]^{1/4} \right\} \quad (23)$$

where the threshold values (as yet unknown) of σ_{fT} and input SNR are used in (23).

The threshold value of input SNR occurs when the IF SNR reaches 10 db. Since the equivalent noise bandwidth of the single-tuned IF filter is $\pi B_{IF}/2$, the threshold input SNR $(\text{SNR})_T$ is expressed by

$$(\text{SNR})_{IF} = (\text{SNR})_T \frac{B_{rf}}{\frac{\pi}{2} B_{IF}} = 10$$

or

$$(\text{SNR})_T = 5\pi B_{IF} B_{rf} \quad (24)$$

where B_{rf} is the noise bandwidth of the RF filter. By substituting in (24) the expression for B_{IF} , (23), we obtain

$$(\text{SNR})_T = \frac{10\pi f_m}{B_{rf}} \left\{ 1 + (2D)^{1/2} \left[\frac{2\pi f_m}{B_{rf}(\text{SNR})_T} \right]^{1/4} \right\}$$

If we define $(\text{SNR})_T B_{rf}/2\pi f_m \triangleq x$, then this equation can be rewritten as $D = \frac{1}{2}(x/5 - 1)^2$, a relationship plotted in Fig. 12. As can be seen from Fig. 12, if

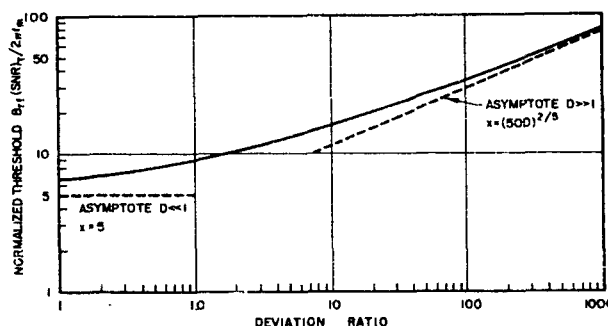


Fig. 12-Normalized threshold input $(\text{SNR})_T$ as a function of the deviation ratio for the frequency-lock discriminator.

$D \gg 1$, then $x \approx (50D)^{2/5}$ is a good approximation. For $D \ll 1$, the normalized SNR x approaches a minimum value of 5. This curve allows the determination of the threshold SNR for an arbitrary RF bandwidth B_{rf} .

A relation of even more interest is the curve of threshold SNR as a function of the deviation ratio D when $B_{rf} = \pi(D + 1)f_m$, its minimum value for low distortion. Then the threshold can be obtained from Fig. 12 by the relation $(\text{SNR})_T = 2x/(D + 1)$, and the result is shown in Fig. 13. As can be seen from the figure, the threshold

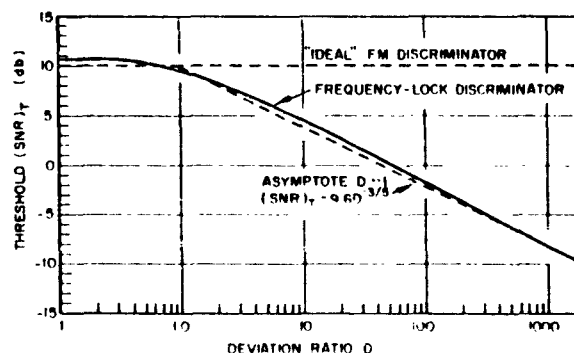


Fig. 13 Threshold $(\text{SNR})_T$ as a function of deviation ratio for the frequency-lock discriminator.

is approximately 10 db for $D \leq 1$, and it decreases approximately as $9.6D^{-3/5}$ for $D > 1$.

Experimental Measurements

An experimental frequency-lock FM discriminator has been constructed in accordance with the theory presented in this paper, and its performance has been measured. The discriminator was designed to demodulate a wide deviation FM signal having the following parameters:

Maximum modulation frequency	$f_m = 1.5 \text{ Kc}$
Peak frequency deviation	$f_d = 1 \text{ Mc}$
Deviation Ratio	$D = 667$

A single-tuned RF filter is used which has a 3-db bandwidth of 2 MC; hence, $B_{rf} = \pi(1 \text{ Mc}) = \pi(D + 1)f_m$.

Figure 14 shows the measured fractional time out of lock for the discriminator as a function of the input SNR. The theoretical threshold value of input SNR based from Fig. 13 is -7.0 db. This prediction agrees well with the measurements which gave a 5 percent time out of lock at input SNR between -6.5 and -7.0 db.

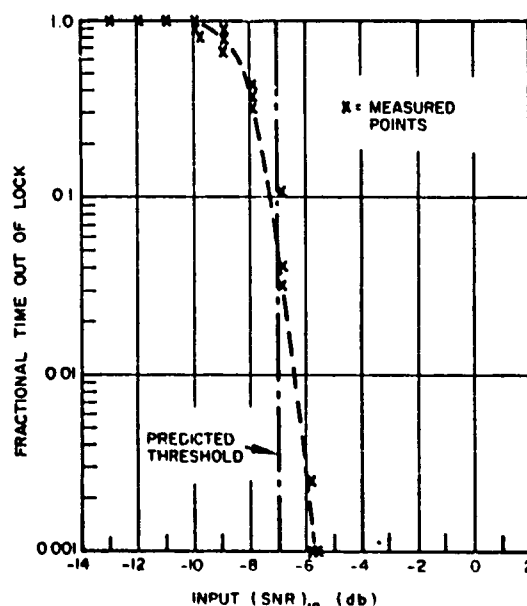


Fig. 14 Measured value of the fractional time out of lock vs. input SNR for the frequency-lock discriminator.

The output SNR for the discriminator was also measured as a function of the input SNR. Because of extraneous circuit noise, hum, etc., the highest output SNR measured was 51 db. The output SNR did not decrease substantially below this value until the input SNR was reduced to a threshold value of between -5.5 and -6.5 db, just as expected from the other measurement. The threshold effect was very prominent. When the input SNR was decreased to -8.5 db, the output SNR decreased sharply to approximately unity.

Maximum Likelihood FM Discriminator

Description

As mentioned in the introduction, the maximum-likelihood FM discriminator can take the same form as the phase-lock loop. However, in that form, it requires a nonrealizable filter. As a particular example, let us consider that the frequency modulation has a low-pass power spectral density $G_f(f)$ of the form

$$G_f(f) = \frac{f_d^2}{4f_m} \quad \text{for } |f| \leq f_m$$

$$= 0 \quad \text{for } |f| > f_m \quad (25)$$

where $f_d/2^{1/2}$ is the rms frequency deviation, and f_m is the maximum modulation frequency. The interfering noise $n(t)$ is considered to be band-limited gaussian white noise having a constant spectral density $G_n(f) = N_0$ in the signal pass band. Under these conditions, the maximum-likelihood FM discriminator can be shown to have the form shown in Fig. 15. The loop filter $F_0(p)$ is nonrealizable because it corresponds to a filter impulse response which is nonzero for $t < 0$.

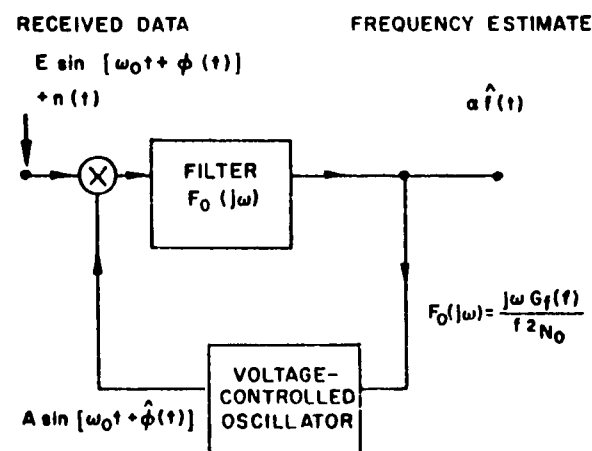


Fig. 15 Maximum-likelihood FM discriminator.

An approximate computer realization* of this estimator is shown in Fig. 16. This computer operates on recorded

*The maximum likelihood estimate can also be obtained by using an a posteriori probability computer. The computer of Fig. 16 is discussed here because it most resembles the form of the phase-lock loop. Both realizations required the use of a complex type of waveform generator.

RECORDED TIME SEGMENT OF RECEIVED DATA

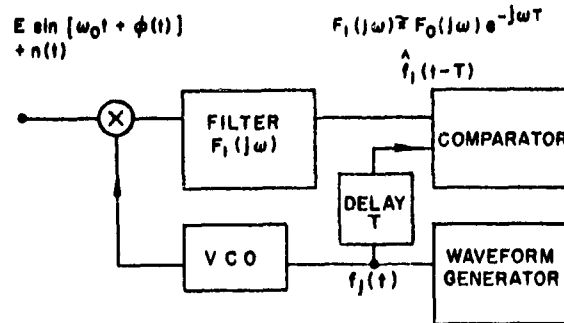


Fig. 16—An approximate computer realization of the maximum-likelihood FM discriminator.

time segments of the received data τ sec. long. It uses a complex type of waveform generator which generates a large number of modulating function which are samples τ sec. long of a waveform having a power spectrum $G_f(f)$ as in (25). Obviously, the number of waveforms to be generated is intolerably large for any but the most simple classes of modulation functions.

The linear filter used in Fig. 16 is an arbitrarily good approximation to the optimum $F_o(p)$ and is realizable because it allows an arbitrary delay of T sec. in the estimation process. The computer operates by comparing the waveforms $f_i(t)$ at the filter output with a delayed version of the test modulating waveform $f_j(t)$. The pair that matches is the maximum likelihood estimate and that $f_j(t)$ is the identical waveform (except for a delay) as is obtained by the estimator of Fig. 15.

To compute the threshold of the maximum likelihood estimator, we make the identical set of calculations as were made for the phase-lock loop, except that the closed-loop transfer function is now of the form

$$\frac{\hat{\Phi}(j\omega)}{\Phi(j\omega)} \triangleq H(j\omega) = \frac{P_s G_f(f)/f^2 N_0}{1 + P_s G_f(f)/f^2 N_0}$$

where $P_s = E^2/2$, the signal power, and $G_f(f)/f^2$ is the power spectral density of the phase modulation $\phi(t)$. Define

$$\lambda = 2f_m/B_{rf}f_d(\text{SNR})_{in}$$

where $(\text{SNR})_{in} = P_s/2B_{rf}N_0$ and B_{rf} is the equivalent noise bandwidth. Then the closed-loop transfer function simplifies to

$$H(j\omega) = \frac{1}{1 + \lambda^2} \quad |f| \leq f_m \quad (26)$$

$$= 0 \quad |f| > f_m$$

The parameters of this filter change with the $(\text{SNR})_{in}$. However, in this paper where threshold is the most important effect, the filter is fixed and optimized for the threshold value of $(\text{SNR})_{in}$. Thus, the filter time constant is chosen $\lambda = 2f_m/B_{rf}(\text{SNR})_T$.

As in the previous computations, threshold is computed on the basis of a sinusoidal frequency modulation, $f(t) = f_d \sin \omega_m t$. The peak transient phase error is $c_t = 2(\lambda/D)/(1 + \lambda/D^2)$. The mean square phase noise error is

$$\sigma_n^2 = \int_{-\infty}^{\infty} G_{\phi_n}(f) |H(j\omega)|^2 df$$

where $G_{\phi_n}(f)$ is the equivalent input phase noise spectrum and $G_{\phi_n} = 1/2 B_{rf}(\text{SNR})_{in}$ in the low-frequency region where $H(j\omega) > 0$. The mean square phase noise can be evaluated as

$$\sigma_n^2 = \frac{1}{B_{rf}(\text{SNR})_{in}} \int_0^{f_m} \frac{df}{(1 + \lambda f^2)^2}$$

$$= \frac{\lambda/4}{1 + \lambda/D^2} + \frac{D\sqrt{\lambda}}{4} \tan^{-1} \sqrt{\lambda}/D$$

Define the threshold to occur when the rms total error, σ_T reaches $1/2$. The equation defining threshold is

$$c_t^2/2 + \sigma_n^2 = (1/2)^2$$

and can be written

$$2\left(\frac{a}{1+a}\right)^2 + \frac{1}{4} \frac{a}{1+a} + \frac{a^{1/2}}{4} \tan^{-1} \frac{1}{\sqrt{a}} = \frac{1}{(2D)^2}$$

where $a \triangleq \lambda/D^2$. This expression can be solved for D as a function of a and then the threshold $(\text{SNR})_{in}$ can be obtained by relating it to a . The result is shown in Fig. 17. For deviation ratios $D > 1$, a good approximation is $(\text{SNR})_T = 1.27/D$.

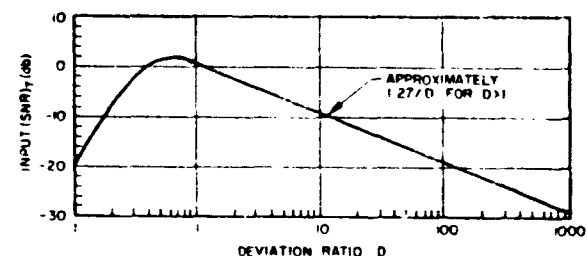


Fig. 17 Threshold $(\text{SNR})_{in}$ as a function of the deviation ratio for the maximum-likelihood estimator

Comparison of the Discriminators

Each of the three types of discriminators has been evaluated on the basis of its threshold value of input SNR for a prescribed type of frequency modulation. When optimum post-detection filtering is used and the discriminators are operating at SNR well above their thresholds, the relationship between the output and input SNR is the same for each type. Thus, it is quite meaningful to compare the performance of the discriminators on the basis of their thresholds.

In Fig. 17, the threshold input SNR is plotted as a function of the deviation ratio for each of the discriminator types. In each example, the equivalent noise bandwidth B_{rf} is assumed to be $\pi(D + 1)f_m$, the minimum value consistent with low distortion of the signal waveform. The +10 db threshold SNR of the conventional or "ideal" discriminator is shown in the figure for comparison.

At small deviation ratios $D < 1$ the thresholds of the two-phase coherent discriminators (phase-lock and maximum-likelihood) actually decrease with decreases in the deviation ratio. The reason for this effect is that the maximum phase excursion caused by the modulation is less than 1 radian in this region. Thus, transient errors are of little significance in determining the threshold, and the noise can be sharply filtered. Of course, for small deviation ratios and low input SNR the output SNR is so small as to be of limited usefulness. Furthermore the deviation caused by drifts in the oscillator center frequencies may have to be taken into account with this type of operation.

The frequency-lock discriminator shows no improvement over the ideal FM discriminator for small deviation ratios. The explanation for this behavior, of course, is that the IF filter has a minimum bandwidth of $B_{IF} = 2f_m$, the same as the RF bandwidth B_{rf} . Hence, the IF filter does not remove any significant amount of noise from the signal.

At deviation ratios $D > 1$ all three types of discriminators provide a reduced threshold input (SNR). As expected, the maximum likelihood FM discriminator has the lowest threshold SNR for all values of the deviation ratio; at high deviation ratios, it has $(SNR)_T = 1.27/D$. The reason for this superiority is that the discriminator is not constrained to real time operation, and the loop filtering can thereby be much more effective. However, as has been stated, the complexity of the waveform generator required of this type of discriminator makes it unsuitable for all but the most elementary classes of modulation functions.

It is shown in Fig. 18 that the phase-lock discriminator is superior in performance to the frequency-lock discriminator for small and moderate values of the deviation ratio ($D < 25$). This behavior can be made plausible by pointing out that the phase-lock discriminator is a phase-coherent device, while the frequency-lock discriminator tracks instantaneous frequency.

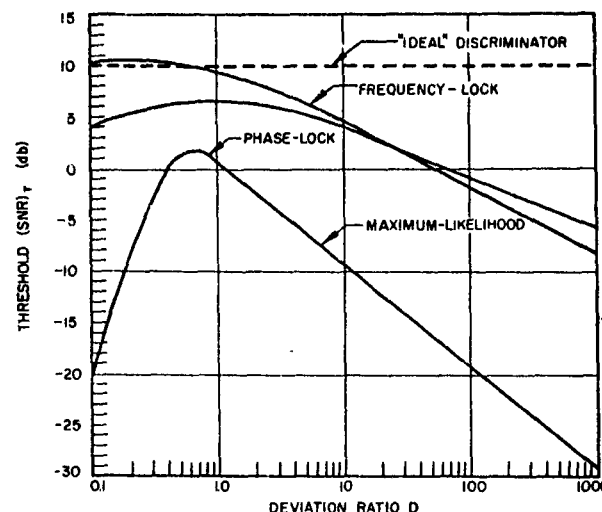


Fig. 18-Comparison of the threshold input SNR for the discriminators as a function of D .

In other words, when the discriminator is locked-on the VCO output is phase coherent with the signal and is noncoherent with the noise. As a result, the noise power in the multiplier output has decreased relative to the signal component by a factor of $1/2$ from its value before the multiplication operation. (This change in noise level is analogous to that found in the synchronous detection of AM signals.) Therefore, the phase coherence of the phase-lock loop can be an important advantage and, evidently, is of dominating importance for deviation ratios $D < 25$. At larger values of the deviation ratio, the frequency-lock discriminator equals or surpasses performance of the phase-lock loop.

This effect can be attributed, at least in part, to a basic difference in the threshold errors of the two discriminators. The threshold phase error for the phase-lock loop is fixed, $|\epsilon| = 1$ is used in the linearized analysis, whereas threshold frequency error for the frequency-lock discriminator is dependent upon an adjustable quantity, the IF bandwidth. The IF bandwidth can, of course, be optimized for a given type of modulation, and in this sense provides a degree of freedom that does not exist in the phase-lock loop. It is interesting to note from (23) that if the IF bandwidth is fixed, perhaps at its optimum value for $D = 1$, then the frequency errors become of dominant importance as D is increased. In this circumstance, the threshold input SNR is proportional to $D^{-1/2}$ for large D , the same relationship as obtained with the phase-lock loop. By optimizing the IF bandwidth as is done in this paper, the threshold input SNR can vary in proportion to $D^{-3/5}$ and, as a result, allows the frequency-lock discriminator to surpass the performance of the phase-lock discriminator at large deviation ratios.

The results of this paper, however, do not permit the conclusion that the frequency-lock technique is superior to the phase-lock technique at large deviation ratios where higher-order transfer functions are allowed.

The analysis becomes substantially more complex, however, as the order of the transfer function is increased to third, or higher order. Furthermore, constraints must then be placed on the closed-loop transfer function in order that the loop filter required be realizable. The absence of poles in the right half-plane of the closed-loop response is insufficient for realizability of the loop filter.

Acknowledgment

The author wishes to acknowledge the valuable contributions made by his associates in the Communications and Controls Research Department, Lockheed Missiles and Space Company. Special acknowledgment goes to R. A. Dye, W. J. Gill, and D. T. Magill who have performed the experimental measurements reported in this paper.

References

1. J. G. Chaffee, "The application of negative feedback to frequency modulation systems," Proc. IRE pp 317-331, May, 1939, also BSTJ, pp 404-437, Jul, 1939.
2. J. R. Carson, "Frequency modulation; theory of the feedback receiving circuits," BSTJ, pp 395-403; Jul, 1939.
3. D. A. Bell, "Reduction of band width in FM receivers," Wireless Engineer, pp 497-502; Nov, 1942.
4. H. W. Bode, "Network Analysis and Feedback Amplifier Design," D. Van Nostrand, New York, pp. 489-493; 1945
5. M. Morita and S. Ito, "High sensitivity receiving system for frequency modulated waves," IRE Convention Record, Part 5, pp 228-237; 1960.
6. T. S. George, "Analysis of synchronizing system for dot-interlaced color television," Proc. IRE, vol. 39, pp 124-131; 1951.
7. G. W. Preston, J. C. Tellier, "The lock-in performance of an AFC circuit," Proc. IRE, pp 249-251; Feb, 1953.
8. D. Richman, "Color carrier reference phase synchronization accuracy in NTSC color TV," Proc IRE, vol. 42, pp 106-133; Jan, 1954.
9. W. J. Gruen, "Theory of AFC synchronization," Proc. IRE, pp 1043-1048; Aug, 1953.
10. R. Jaffe, E. Reichtin, "Design and performance of phase-locked circuits capable of near optimum performance over a wide range of input signal-to-noise ratios," IRE Trans. vol. IT-1, pp 66-72, Mar, 1955.
11. C. F. Gilchrist, "Application of phase-locked loop to telemetry as a discriminator or tracking filter," IRE Trans. vol. T & RC-4, pp 20-35; Jun, 1958.
12. G. W. Preston, "Basic theory of locked-oscillators in tracking FM signals," IRE Trans. vol. SET-5, pp 30-33; Mar, 1959.
13. S. G. Margolis, "The response of a phase-locked loop to a sinusoid plus noise," IRE Transactions PGIT, pp 136-143; Jun, 1957.
14. C. S. Weaver, "A new approach to the linear design and analysis of phase-locked loops," IRE Trans. vol. SET-5, pp 166-179; Dec, 1959.
15. F. W. Lehan, R. J. Parks, "Optimum demodulation," IRE Convention Record, Part 8, PGIT, pp 101-103; 1953.
16. J. J. Spilker, Jr., "Statistical design and evaluation of demodulation systems for angle modulated sinusoids," Lockheed Missiles and Space Division, LMSD-48496, Sunnyvale, Calif., Apr, 1959.
17. W. B. Davenport, "Signal-to-noise ratios in band-pass limiters," Mass. Institute of Tech., Res. Lab. of Electronics, TR No. 234, pp 11-17; May, 1952.
18. S. O. Rice, "Statistical properties of a sine wave plus random noise," Bell System Tech. Journal, vol. 27, pp 109-157; Jan, 1948.
19. E. J. Baghdady, "Theory of low-distortion reproduction of FM signals in linear systems," IRE Trans., Circuit Theory, pp 202-214; Sep, 1958.
20. F. L. H. M. Stumpers, "Theory of frequency modulation noise," Proc. IRE., pp 1081-1092; Sep, 1948.

Appendix A

Optimization of the Phase-Lock Loop Transfer Function

In this appendix, the phase lock loop, closed-loop transfer function is assumed to have the form,

$$H(p) = \frac{1 + ap}{1 + b_1 p + b_2 p^2}$$

The parameters of this function are to be optimized so as to minimize the rms total phase error in the modulation estimate. From (6) the transient phase error can be evaluated as

$$\epsilon_t = \frac{f_d}{f_a} |1 - H(j\omega_a)| = \frac{f_d}{f_a} \left[\frac{\omega_a^2 (b_1 - a)^2 + b_2^2 \omega_a^4}{(1 - \omega_a^2 b_2)^2 + \omega_a^2 b_1^2} \right]^{1/2}$$

The transient error is largest for $f_a = f_m$ the maximum modulation frequency, and this value will be assumed henceforth. For large deviation ratios and $|\epsilon| \ll 1$, this expression can be simplified to

$$1 - \left(\frac{\epsilon_t}{D} \right)^2 = \frac{\omega_m^2 (b_1 - a)^2 + b_2^2 \omega_m^4}{D^2} \quad (27)$$

These conditions of large deviation ratio and $|\epsilon_t|$ below the threshold value are of major importance in this paper, and it is for these conditions that the closed-loop transfer function is optimized.

The mean square phase noise error can be computed from (7) and found to be

$$\sigma_n^2 = \int_{-\infty}^{\infty} G_{\phi_n}(f) |H(j\omega)|^2 df = \frac{G_{\phi_n}(0)}{b_1} \left[1 + \frac{a^2}{b_2^2} \right] \quad (28)$$

where $G_{\phi_n}(f) \cong G_{\phi_n}(0)$ for deviation ratios $D > 1$. It can be shown from (2) that the power-spectral density

$$G_{\phi_n}(0) = \left[2B_{rf}(SNR)_{in} \right]^{-1},$$

where B_{rf} is the equivalent noise bandwidth.

Thus, the total mean square phase error is of the form

$$\sigma_T^2 = \epsilon_t^2/2 + \sigma_n^2$$

and has the representation

$$\sigma_T^2 = \frac{D^2}{2} \left[(b_1 - a)^2 \omega_m^2 + b_2^2 \omega_m^4 \right] + \frac{\omega_m}{2B_{rf}(SNR)_{in}} \left[\frac{b_2 + a^2}{\omega_m b_1 b_2} \right]$$

Define the quantities

$$\alpha = a\omega_m, \beta_1 = b_1\omega_m, \beta_2 = b_2\omega_m^2, B = \pi\omega_m/2B_{rf}(SNR)_{in}$$

Then this expression can be rewritten

$$\sigma_T^2 = \frac{D^2}{2} \left[(\beta_1 - \alpha)^2 + \beta_2^2 \right] + B \left(\frac{\beta_1 + \alpha^2}{\beta_1 \beta_2} \right)$$

It is easily seen that the optimum value of α is bounded by $0 < \alpha < \beta_1$. A good approximation to the approximate minimum of σ_T^2 is obtained with $\alpha = \beta_1$, and this value of α will be used. The optimum value of β_2 is then β_1^2 , and the resulting close-loop transfer function is

$$H(p) = \frac{1 + ap}{1 + ap + a^2 p^2}$$

The peak transient error (27) now has the value

$$\epsilon_t = D a^2 \omega_m^2$$

and the mean square noise error (28) is

$$\sigma_n^2 = \left[2 a B_{rf}(SNR)_{in} \right]^{-1}$$

Hence the input SNR is related to the transient and noise phase errors by the equation

$$(SNR) = \frac{\pi f_m}{B_{rf}(\epsilon_t/D)^{1/2} \sigma_n^2}$$

This result is the same as (10) in the text.

Appendix B

Optimization of the Frequency-Lock Transfer Function

The closed-loop transfer function of the frequency-lock discriminator is considered to have the form

$$H(p) = \frac{1}{1 + b_1 p + b_2 p^2}$$

When sinusoidal frequency modulation at frequency f_a is applied to the signal, the peak transient frequency error can be found from (18) to be

$$f_{tp} = f_d |1 - H(j\omega_a)|$$

$$= f_d \left[\frac{b_1^2 \omega_a^2 + b_2^2 \omega_a^4}{(1 - b_2 \omega_a^2)^2 + b_1^2 \omega_a^2} \right]$$

The maximum transient error occurs for $f_a = f_m$, the maximum modulation frequency, and this modulation frequency will be used. For large ratios of the frequency deviation to the transient frequency error, the situation of greatest interest, this expression reduces to

$$(f_{tp}/f_d) \cong b_1^2 \omega_m^2 + b_2^2 \omega_m^4$$

The mean square value of the frequency noise error can be computed from (19) and has the value

$$\sigma_{f_n}^2 = \int_{-\infty}^{\infty} G_{f_{ne}}(f) |H(j\omega)|^2 df$$

where $G_{f_{ne}}(f)$ is the power spectrum of the equivalent frequency noise input,

$$G_{f_{ne}}(f) = f^2 / \left[2B_{rf}(SNR)_{in} \right]$$

Thus $G_{f_{ne}}(f)$ is a parabolic power spectrum. Evaluation of this integral gives the result

$$\sigma_{f_n}^2 = \left[16\pi^2 B_{rf}(\text{SNR})_{\ln} b_1 b_2 \right]^{-1}$$

If we now define

$$\beta_1 = b_1 \omega_m, \beta_2 = b_2 \omega_m^2, B = \pi f_m^3 / \left[2B_{rf}(\text{SNR})_{\ln} \right],$$

then the mean square total frequency error,

$$\sigma_{fT}^2 = f_{tp}^2/2 + \sigma_{f_n}^2$$

is

$$\sigma_{f_n}^2 = \frac{f_d^2}{2} (\beta_1^2 + \beta_2^2) + \frac{B}{\beta_1 \beta_2}$$

This expression is symmetrical with respect to β_1 and β_2 and is minimized by setting $\beta_1 = \beta_2$, i.e., $b_2 = b_1/\omega_m$. Hence the optimum form of $H(p)$ is

$$H(p) = \frac{1}{1 + b_1 p^2 + b_1 p^2 / \omega_m}$$

The resulting mean square error is then

$$\sigma_{fT}^2 = f_d^2 \beta_1^2 + B/\beta_1^2$$

This error is minimized if

$$\beta_1 = (B/f_d^2)^{1/4},$$

and thus the minimum mean square frequency error is

$$\sigma_{fT}^2 = 2f_d B^{1/2} \left[\frac{2\pi f_m}{B_{rf}(\text{SNR})_{\ln}} \right]^{1/2}$$

The square root of this quantity, σ_{fT} , gives (22) of the text.